# Randomness Extraction Fails for Finite-State Dimension

**Subin Pulari**
National Research University Higher School of Economics Moscow, Russia

**Akhil S**
Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India

## Abstract

Finite-state dimension, introduced as a finite-state analogue of Hausdorff dimension, quantifies the lower asymptotic density of information in an infinite sequence as perceived by finite-state automata. It admits several equivalent formulations; two particularly useful are via finite-state gambling strategies and via the optimal asymptotic compression ratio achieved by information-lossless finite-state compressors.

Normal sequences represent the highest level of algorithmic randomness visible to finite automata, and are exactly those sequences having finite-state dimension equal to 1. This motivates a bounded-memory notion of randomness extraction: can a finite-state transducer, reading a single sequence streamingly, extract a normal output from a single input source? More modestly, can it always transform the input into an output of strictly higher finite-state dimension? Finite-state transducers can perform surprisingly effective one-pass transformations: even with constant memory they can implement variable-length coding schemes including Shannon–Fano coding, remove local redundancy, and increase the apparent randomness rate on many structured or stochastic inputs.

We show randomness extraction using transducers is impossible in a strong, explicit form. For every rational $s \in (0,1)$, we construct a near linear-time computable binary sequence $X$ with $\dim_{\mathrm{FS}}(X) = s$ such that for every finite-state transducer $T$, the output satisfies $\dim_{\mathrm{FS}}(T(X)) \leq s$. Thus, for these sequences, finite-state transduction cannot extract normality – indeed it cannot even improve finite-state dimension. Our proof proceeds by a structural analysis of finite-state transducers together with a dimension-preserving diagonal construction that, for each target $s$, builds a sequence whose organization defeats every such transducer's attempt to concentrate randomness. The result is a finite-state analogue of Miller's non-extractability phenomenon for effective dimension, but its proof relies on substantially different techniques, tailored to the finite-state setting.

Furthermore, we show that the impossibility persists even with multiple independent input streams. We treat two notions of independence: (i) Kolmogorov-complexity–based independence (via joint prefix complexity), and (ii) a finite-state notion of relative independence, formulated via relative finite-state dimension. By sharp contrast with the effective-dimension setting – where two independent sources suffice for a uniform effective procedure that boosts randomness rate arbitrarily close to 1 – we show that finite-state dimension exhibits no comparable multi-source extraction phenomenon. Specifically, for every rational $s \in (0,1)$ and every fixed $k \geq 2$, there exist $k$ independent sources, each of finite-state dimension $s$, such that for every $k$-input finite-state transducer $T$, the output satisfies $\dim_{\mathrm{FS}}(T(X_1, \ldots, X_k)) \leq s$. Thus, even independent streams do not allow bounded-memory transduction to output a normal sequence or to increase finite-state dimension.

**2012 ACM Subject Classification** Mathematics of computing → Information theory; Theory of computation → Formal languages and automata theory; Theory of computation → Transducers

**Keywords and phrases** Finite-state dimension, normal numbers, randomness extraction, finite-state transducers

**Digital Object Identifier** 10.4230/LIPIcs.LICS.2026.78

**Funding** *Subin Pulari*: supported by the Basic Research Program at HSE University and this paper was prepared in the framework of this program.
*Akhil S*: supported by the Research-I Foundation of the Department of Computer Science and Engineering, IIT Kanpur, for research and conference travel.





**Acknowledgements** The authors would like to thank Jun Le Goh for proposing the question studied in this paper. They also thank Laurent Bienvenu, Elvira Mayordomo and Satyadev Nandakumar for helpful discussions.

## 1 Introduction

Randomness can be understood as the absence of exploitable structure. In algorithmic information theory this idea is made precise by the equivalence between randomness and incompressibility: Martin–Löf random sequences are exactly those whose prefixes have Kolmogorov complexity essentially as large as their length. This viewpoint naturally suggests *extraction* questions: given a source that is only partially random, can an effective transformation reorganize its information so that the output looks more random?

A quantitative way to ask this is via *effective dimension* [13], which measures the asymptotic density of algorithmic information in a sequence. Here $K(\sigma)$ denotes the plain Kolmogorov complexity of a finite string $\sigma$, i.e., the length of the shortest program that outputs $\sigma$ on a fixed universal machine.[1] For a binary sequence $X$, the constructive (Hausdorff) dimension

$$\dim(X) = \liminf_{n\to\infty} \frac{K(X\upharpoonright n)}{n}$$

captures the lower asymptotic information rate of its prefixes, while the constructive strong dimension

$$\mathrm{Dim}(X) = \limsup_{n\to\infty} \frac{K(X\upharpoonright n)}{n}$$

captures the corresponding upper rate [15]. In this language, dimension extraction asks whether one can increase the constructive dimension of a sequence, ideally to 1, by computable transformations. In the single-source setting this can fail in the strongest possible way: Miller [16] constructed a $\Delta^0_2$ sequence of effective dimension $1/2$ whose Turing degree contains no sequence of higher effective dimension. On the other hand, the obstruction is genuinely single-source: under suitable algorithmic independence assumptions, Zimand [22] showed that two positive-dimension sources can be combined by an effective procedure to yield dimension 1. A related line of work studied extraction at the level of Kolmogorov complexity: Fortnow et al. [8] showed that such extraction results yield zero–one laws for strong dimension in both resource-bounded and constructive settings.

This paper studies the analogous extraction question at the opposite end of the computational spectrum, where "randomness" is judged only by *bounded memory*. *Finite-state dimension*, introduced by Dai, Lathrop, Lutz, and Mayordomo [6], assigns to each infinite sequence a number in $[0, 1]$ that can be read as the *randomness rate visible to finite automata*. It is the finite-state analogue of effective dimension, and it admits robust equivalent characterizations via finite-state gamblers ($s$-gales), block entropy rates [3, 1], finite-state prediction [9], automatic Kolmogorov complexity [11] and finite-state (streaming) compression [3]. The corresponding upper analogue, the *strong finite-state dimension* $\mathrm{Dim}_{\mathrm{FS}}$ [1], is obtained by replacing lower asymptotic rates by upper asymptotic rates (e.g., via limsup versions of the same characterizations). In this setting, the classical notion of Borel normality plays the

[1] Although $K$ is often used to denote prefix-free Kolmogorov complexity, we follow Zimand [22] in using $K$ to denote plain Kolmogorov complexity throughout the paper.

role of full randomness: a sequence is normal iff every fixed block appears with the limiting frequency prescribed by the uniform measure, and normality coincides with having $\dim_{\mathrm{FS}} = 1$ (equivalently, no finite-state gambler can win on it [19]).

In the finite-state setting, extraction is naturally carried out by finite-state transducers, which act as online, constant-memory compressors. A finite-state transducer (FST) processes an input stream in one pass and emits a variable-length output word per symbol. This variable-length output capability already suffices to implement standard block codes (e.g., Shannon–Fano or Huffman) in a single pass, letting an FST systematically strip local redundancy and lock onto simple structured regularities using only constant memory. When the transducer is information-lossless (so that finite prefixes can be recovered from the output together with the final state), it forms a classical model of streaming compression. For an information-lossless finite-state compressor $C$ and a sequence $S$, define the lower and upper compression ratios

$$\rho_C^-(S) = \liminf_{n\to\infty} \frac{|C(S\upharpoonright n)|}{n}, \qquad \rho_C^+(S) = \limsup_{n\to\infty} \frac{|C(S\upharpoonright n)|}{n}.$$

One of the most useful perspectives on $\dim_{\mathrm{FS}}$ and $\mathrm{Dim}_{\mathrm{FS}}$ is that they coincide with the best achievable asymptotic compression ratios :

$$\dim_{\mathrm{FS}}(S) = \inf_C \rho_C^-(S), \qquad \mathrm{Dim}_{\mathrm{FS}}(S) = \inf_C \rho_C^+(S),$$

where $C$ ranges over information-lossless finite-state compressors [6, 3]. From this viewpoint, "extracting randomness" corresponds to compressing away finite-state-detectable structure so that the remaining stream has higher finite-state dimension.

This compression viewpoint already yields nontrivial one-source phenomena. Doty [7] showed that information-lossless finite-state transducers can partially boost finite-state dimension: if a source has $\dim_{\mathrm{FS}}(S) = d$ and $\mathrm{Dim}_{\mathrm{FS}}(S) = D > 0$, then (nonuniformly, depending on $S$) one can produce an output whose finite-state dimension is at least $d/D - \varepsilon$. Thus, when $D < 1$, bounded memory can increase the lower randomness rate by exploiting the *absence* of fluctuations in information density.

The central question motivating this work is whether bounded-memory extraction can be guaranteed in general.

*Can bounded-memory transducers always increase the finite-state randomness rate of a source? A separate extremal question asks whether every sequence admits a finite-state transducer whose output is normal.*

We answer this in the negative, and in a way that sharply contrasts the effective multi-source picture. Our first main theorem is a finite-state analogue of Miller's non-extraction phenomenon.

**Single-source non-extraction (informal).** *For every rational $s \in (0, 1)$ there exists a binary sequence $X$ with $\dim_{\mathrm{FS}}(X) = s$ such that for every finite-state transducer $T$,*

$$\dim_{\mathrm{FS}}(T(X)) \le s.$$

*In particular, no bounded-memory transduction can extract normality from $X$.*

At a technical level, this statement goes beyond limitations of information-lossless compression: it rules out *all* finite-state transducers, including those that discard information. It also aligns with the compression-based viewpoint above: the constructed sequences satisfy $\mathrm{Dim}_{\mathrm{FS}}(X) = 1$, so general $d/D$-type lower bounds do not force any increase beyond $s$, and we show that no finite-state transducer can improve upon this.

We also study the multi-source setting, where a transducer reads $k$ input streams *in lockstep* (symbol-by-symbol) and produces a single output stream. In the effective setting, multi-source extraction becomes possible under suitable Kolmogorov-complexity independence hypotheses [22]. It is therefore natural to ask whether multiple sources can help bounded-memory transducers extract higher *finite-state* randomness rates. We address this question under two independence hypotheses that play different conceptual roles.

We consider two notions of independence for a pair of input streams $X, Y \in \{0,1\}^\infty$. First, we use the standard *Kolmogorov-complexity independence* hypothesis (as in, e.g., [22]), requiring that for all $n, m \in \mathbb{N}$,

$$K\big(X\upharpoonright n,\ Y\upharpoonright m\big)\ \geq\ K(X\upharpoonright n) + K(Y\upharpoonright m)\ -\ O(\log n + \log m).$$

Equivalently, knowing a prefix of one stream does not substantially shorten the shortest description of a prefix of the other (beyond logarithmic slack).

Second, we use a genuinely finite-state notion based on *finite-state relative dimension* in the sense of [17]. Informally, one measures the disjoint-block conditional entropy rate of one stream given another stream revealed in lockstep, and defines $\dim_{\mathrm{FS}}^{Y}(X)$ accordingly. We call $X$ and $Y$ *relatively independent* if conditioning does not reduce finite-state dimension,

$$\dim_{\mathrm{FS}}^{Y}(X) = \dim_{\mathrm{FS}}(X) \qquad \text{and} \qquad \dim_{\mathrm{FS}}^{X}(Y) = \dim_{\mathrm{FS}}(Y),$$

with the natural extension to $k$ sources via mutual relative independence.

For each of these hypotheses, we show that multi-source extraction still fails.

**Multi-source non-extraction (informal).** *Fix $k \geq 2$ and a rational $s \in (0,1)$. For each of the two independence notions above, there exist $k$ sequences $X_1, \ldots, X_k$ with $\dim_{\mathrm{FS}}(X_i) = s$ and $\mathrm{Dim}_{\mathrm{FS}}(X_i) = 1$ for every $i$, satisfying that notion of independence, such that every $k$-input finite-state transducer $T$ satisfies*

$$\dim_{\mathrm{FS}}\big(T(X_1, \ldots, X_k)\big) \leq s.$$

In particular, bounded memory cannot extract normality, and cannot even increase finite-state dimension, from these sources.

Beyond the qualitative separation from the classical setting, our results are constructive and efficient. The witnessing sequences are computable in nearly linear time: the first $n$ bits of each $X$ (and of each $X_i$ in the multi-source construction) can be generated in time $n^{1+o(1)}$.

Conceptually, the proofs exploit graph-theoretic and combinatorial constraints specific to finite machines. Whereas Miller-style arguments rely on the fine structure of Turing reductions and Kolmogorov complexity, here the obstruction comes from the finite directed-graph structure of a transducer: its runs are walks in a bounded state graph. We design a sequence that has sufficiently large random (normal) phases to attain the desired finite state dimension. At the same time, we exploit the directed-graph structure of a transducer to fix controlled, predictable phases that steer the walk into prescribed strongly connected components with long enough stretch of controlled cycling. This ensures that a fixed fraction of the output stream has low block entropy, permanently capping $\dim_{\mathrm{FS}}$.

Together, these results identify a structural limitation of bounded-memory transduction. The obstruction is not merely a single-source phenomenon: it persists even in the multi-source setting, and even when the sources satisfy either Kolmogorov-complexity independence or finite-state relative independence. This gives a sharp conceptual separation from the

effective-dimension setting. There, suitable independence hypotheses can make multi-source extraction possible, while here even multiple independent finite-state sources need not allow any increase in finite-state dimension. Thus finite-state dimension is markedly less amenable to extraction by bounded-memory transduction, even when several independent streams are available.

## 2 Preliminaries

### Strings, sequences, and indexing

Let $\Sigma$ be a finite alphabet with $|\Sigma| \geq 2$. Write $\Sigma^*$ for the set of finite strings over $\Sigma$ (including the empty string $\lambda$) and $\Sigma^\infty$ for the set of infinite sequences. For $w \in \Sigma^*$, let $|w|$ denote its length. For $w \in \Sigma^*$ and integers $0 \leq i \leq j \leq |w|$, write

$$w[i:j] \;:=\; w[i]\,w[i+1]\cdots w[j-1]$$

for the substring from position $i$ (inclusive) to $j$ (exclusive), and $w[:n] = w[0:n]$ for the length-$n$ prefix. For $X \in \Sigma^\infty$ we use the same notations: $X[:n]$ is the length-$n$ prefix and, for $0 \leq i < j$, $X[i:j]$ is the finite block $X[i]\cdots X[j-1]$. When a slice has the form $X[i:i+\ell]$, it denotes the length-$\ell$ block starting at position $i$. Concatenation of finite strings is denoted by juxtaposition or by $\cdot$ when clarity is needed.

We use $\lfloor\cdot\rfloor$ for the floor function and $a \bmod m$ for the residue of $a$ modulo $m$ in $\{0,1,\ldots,m-1\}$. For a predicate $P$, $\mathbf{1}[P] \in \{0,1\}$ denotes its indicator.

### Basic probability and information

All logarithms are base 2 unless explicitly stated otherwise. For a finite probability distribution $p$ on a finite set $\Omega$, its Shannon entropy [5] is

$$H(p) \;:=\; -\sum_{\omega\in\Omega} p(\omega)\log p(\omega), \text{ with the convention } 0\log 0 := 0.$$

For a joint distribution $p_{X,Y}$ on $\Omega_X \times \Omega_Y$, the conditional entropy is

$$H(X \mid Y) \;:=\; \sum_{y\in\Omega_Y} p_Y(y)\, H\big(p_{X|Y=y}\big).$$

The binary entropy function is $h : [0,1] \to [0,1]$ given by

$$h(\alpha) \;:=\; -\alpha\log\alpha - (1-\alpha)\log(1-\alpha) \qquad (\text{with } 0\log 0 := 0).$$

For distributions $p, q$ on the same finite space $\Omega$, the total variation distance is

$$\|p-q\|_{\mathrm{TV}} \;:=\; \frac{1}{2}\sum_{\omega\in\Omega}|p(\omega)-q(\omega)| \;=\; \frac{1}{2}\|p-q\|_1.$$

### Directed graphs

We will refer to standard graph-theoretic notions when analyzing finite-state devices. A directed graph is a pair $G = (V, E)$ with $E \subseteq V \times V$. A (directed) walk is a sequence $(v_0,\ldots,v_t)$ such that $(v_{i-1}, v_i) \in E$ for all $i$. A (directed) path is a walk with no repeated vertices. A directed cycle is a walk with $t \geq 1$ and $v_0 = v_t$; it is *simple* if $v_0,\ldots,v_{t-1}$ are all distinct. Vertices $u, v$ are *mutually reachable* if there are directed paths $u \rightsquigarrow v$ and $v \rightsquigarrow u$.

A *strongly connected component* (SCC) is a maximal subset $C \subseteq V$ such that every pair of vertices in $C$ is mutually reachable. The SCCs form a partition of $V$, and contracting each SCC to a single node yields the *condensation graph*, which is acyclic. An SCC $C$ is *terminal* if no edge leaves it, i.e., there do not exist $u \in C,\ v \in V \setminus C$ with $(u, v) \in E$, equivalently, $C$ is a sink node in the condensation graph.

## 2.1 Finite-state transducers and induced outputs

We use finite-state transducers as our model of bounded-memory stream processing.

▶ **Definition 1** ($k$-input finite-state transducer)**.** *Fix $k \in \mathbb{N}$. A (deterministic) $k$-input finite-state transducer ($k$-FST) is a tuple*

$$T = (Q, \Sigma^k, \Gamma, q_0, \delta, \tau),$$

*where $Q$ is a finite set of states, $\Sigma$ is the per-track input alphabet, $\Gamma$ is the output alphabet, $q_0 \in Q$ is the start state, $\delta : Q \times \Sigma^k \to Q$ is the transition function, and $\tau : Q \times \Sigma^k \to \Gamma^*$ is the output function. When $k = 1$ we simply call $T$ an FST.*

For each transducer $T = (Q, \dots)$ with $Q = \{1, \dots, |Q|\}$, we use the natural order $1 < 2 < \cdots < |Q|$.

▶ **Definition 2** (Extended transition/output)**.** *Define $\widehat{\delta} : Q \times (\Sigma^k)^* \to Q$ and $\widehat{\tau} : Q \times (\Sigma^k)^* \to \Gamma^*$ by $\widehat{\delta}(q, \lambda) = q$, $\widehat{\tau}(q, \lambda) = \lambda$, and for $w \in (\Sigma^k)^*$, $a \in \Sigma^k$,*

$$\widehat{\delta}(q, wa) = \delta(\widehat{\delta}(q, w), a), \qquad \widehat{\tau}(q, wa) = \widehat{\tau}(q, w) \cdot \tau(\widehat{\delta}(q, w), a).$$

*For $Z \in (\Sigma^k)^\infty$, define $T(Z)$ (when the limit exists) by the unique infinite word $Y$ such that*

$$Y \restriction |\widehat{\tau}(q_0, Z \restriction n)| = \widehat{\tau}(q_0, Z \restriction n) \quad \text{for all } n.$$

Except in the final section, we always consider the single-input case $k = 1$.

### Transducers as directed graphs

Let $T = (Q, \{0,1\}, \{0,1\}, q_0, \delta, \tau)$ be a binary-input, binary-output finite-state transducer. When analyzing $T$ graph-theoretically, we use the following *labeled directed multigraph* representation. Define the (labeled) edge set

$$E_T \ := \ \{(q, a, \delta(q, a)) : q \in Q,\ a \in \{0,1\}\}.$$

We think of each triple $(q, a, q') \in E_T$ as a directed edge from $q$ to $q'$ labeled by the input symbol $a$ and carrying the output word $\tau(q, a)$. In particular, *parallel edges are allowed*: it may happen that $\delta(q, 0) = \delta(q, 1)$, yielding two distinct edges with the same endpoints but different labels/outputs.

When we speak of an SCC of $T$, we mean an SCC of the *underlying* directed graph on vertex set $Q$ obtained from $E_T$ by forgetting the label $a$ (and identifying parallel edges if desired), since SCC membership depends only on reachability. However, whenever we refer to a directed walk/cycle in $T$ together with its *input label* and *output word*, we mean a walk/cycle in the labeled edge set $E_T$, so that the input label and output are uniquely determined by the chosen edge sequence.

## 2.2 Block frequencies, entropy rates, normality and finite-state dimension

Finite-state dimension can be characterized in several equivalent ways. In this paper we use the *block-entropy characterization* in terms of empirical block frequencies (see [3]). Both *disjoint* block parsing and *sliding-window* parsing appear in the literature; we begin by recording both frequency notions.

▶ **Definition 3** (Disjoint and sliding $\ell$-block frequencies). *Fix $\ell \in \mathbb{N}$. For $w \in \Sigma^*$ with $|w| \geq \ell$ and $u \in \Sigma^\ell$, define the* disjoint $\ell$*-block frequency*

$$P_\ell^{\mathrm{d}}(u,w) \;:=\; \frac{1}{k}\big|\{0 \leq i < k :\; w[i\ell : (i+1)\ell] = u\}\big|, \qquad k = \Big\lfloor \frac{|w|}{\ell} \Big\rfloor,$$

*and the* sliding-window $\ell$*-block frequency*

$$P_\ell^{\mathrm{s}}(u,w) \;:=\; \frac{1}{|w|-\ell+1}\big|\{0 \leq i \leq |w|-\ell :\; w[i : i+\ell] = u\}\big|.$$

*Unless explicitly stated otherwise, we write $P_\ell(u,w)$ (or simply $P(u,w)$ when $\ell$ is clear) for $P_\ell^{\mathrm{d}}(u,w)$.*

We now define the normalized disjoint $\ell$-block entropy rate and the induced finite-state dimension. This is the characterization we will use throughout [3].

▶ **Definition 4** (Disjoint $\ell$-block entropy rate and finite-state dimension [3]). *For $w \in \Sigma^*$ with $|w| \geq \ell$, define*

$$H_\ell(w) = -\frac{1}{\ell \log |\Sigma|} \sum_{u \in \Sigma^\ell} P_\ell(u,w) \log P_\ell(u,w),$$

*with the convention $0 \log 0 = 0$. For $X \in \Sigma^\infty$, define*

$$H_\ell(X) = \liminf_{k\to\infty} H_\ell(X \upharpoonright k\ell), \qquad \dim_{\mathrm{FS}}(X) = \inf_{\ell \geq 1} H_\ell(X).$$

The *finite-state strong dimension* of $X$ is $\mathrm{Dim}_{\mathrm{FS}}(X) = \inf_{\ell\geq 1} \limsup_{k\to\infty} H_\ell(X \upharpoonright k\ell)$ [3, 1]. Normality is the classical notion of having the correct limiting block frequencies at every length. It is important for us because, over the binary alphabet, normality coincides exactly with having finite-state dimension 1.

▶ **Definition 5** (Normality). *A binary sequence $R \in \{0,1\}^\infty$ is* normal *if for every $\ell \in \mathbb{N}$ and every $u \in \{0,1\}^\ell$,*

$$\lim_{k\to\infty} P_\ell\big(u,\, R \upharpoonright k\ell\big) = 2^{-\ell}.$$

▶ **Theorem 6** ([3]). *A binary sequence $R \in \{0,1\}^\infty$ is normal if and only if $\dim_{\mathrm{FS}}(R) = 1$.*

▶ Remark. It is known that replacing disjoint $\ell$-block frequencies by sliding-window $\ell$-block frequencies leaves $\dim_{\mathrm{FS}}$ unchanged [11]. Similarly, the definition of normality is unchanged if disjoint block frequencies are replaced by sliding-window block frequencies.

## 2.3 Kolmogorov complexity and independence

Following the notation and conventions of [22], we use Kolmogorov complexity to formalize independence between infinite sequences (via their finite prefixes).

### Kolmogorov complexity

We use plain Kolmogorov complexity $K(\cdot)$ on $\{0,1\}^*$ with respect to a fixed optimal machine. For a sequence $X \in \{0,1\}^\infty$, $X \upharpoonright n$ denotes its length-$n$ prefix. For finite strings $U, V$, we write $K(U\ V)$ for $K(\langle U, V\rangle)$ under a fixed computable pairing function (any two standard pairings agree up to an additive $O(\log|U| + \log|V|)$ term). We now recall the independence notion from [22].

▶ **Definition 7** ([22]). *Two sequences $X, Y \in \{0,1\}^\infty$ are* independent *if for all $n, m \in \mathbb{N}$,*

$$K\big(X \upharpoonright n \ \ Y \upharpoonright m\big) \ \geq \ K(X \upharpoonright n) + K(Y \upharpoonright m) - O(\log n + \log m).$$

*Equivalently (up to the same additive $O(\log n + \log m)$ term), for all $n, m \in \mathbb{N}$,*

$$K(X \upharpoonright n \mid Y \upharpoonright m) \ \geq \ K(X \upharpoonright n) - O(\log n + \log m),$$
$$K(Y \upharpoonright m \mid X \upharpoonright n) \ \geq \ K(Y \upharpoonright m) - O(\log n + \log m).$$

Since we also consider a multi-source model, we use the natural $k$-ary generalization to tuples of sequences.

▶ **Definition 8** ($k$-way independence). *A $k$-tuple $(X_1, \ldots, X_k) \in (\{0,1\}^\infty)^k$ is* independent *if for all $n_1, \ldots, n_k \in \mathbb{N}$,*

$$K\big(X_1 \upharpoonright n_1 \ \cdots \ X_k \upharpoonright n_k\big) \ \geq \ \sum_{j=1}^{k} K(X_j \upharpoonright n_j) \ - \ O\Big(\sum_{j=1}^{k} \log n_j\Big).$$

We will use the following standard inequalities [22, 12].

▶ **Lemma 9.** *For all finite strings $U, V$:*
1. *If $U$ is computable from $(V, W)$, then $K(U) \leq K(V) + |W| + O(1)$.*
2. *(Chain rule) $K(UV) \geq K(U) + K(V \mid U) - O(\log|U| + \log|V|)$.*

Finally, we record that independent normal sources exist.

▶ **Lemma 10.** *There exist sequences $A, B \in \{0,1\}^\infty$ such that (i) $A$ and $B$ are normal, and (ii) $A$ and $B$ are independent in the sense of Definition 7. More generally, for every $k \geq 2$ there exist $A_1, \ldots, A_k \in \{0,1\}^\infty$ that are normal and $k$-way independent in the sense of Definition 8.*

A standard way to obtain such sources is to iterate relativized Martin–Löf randomness ($A_1$ random, $A_2$ random relative to $A_1$, and so on), and then translate the resulting conditional bounds to plain complexity with only logarithmic overhead.

## 2.4 Finite-state relative dimension and relative independence

Following [17], we define a finite-state notion of relative independence via finite-state *relative* dimension, using its disjoint-block conditional-entropy characterization.

We begin by defining the empirical disjoint $\ell$-block joint and conditional distributions.

▶ **Definition 11** (Disjoint $\ell$-block distributions [17]). *Fix $\ell \in \mathbb{N}$ and $k \in \mathbb{N}$. Let $x \in \{0,1\}^{k\ell}$ and let $y \in \Delta^{k\ell}$ for some finite alphabet $\Delta$. For $u \in \{0,1\}^\ell$ and $v \in \Delta^\ell$, define*

$$P_{x,y}(u,v) = \frac{1}{k}\big|\{0 \leq i < k : \ x[i\ell : (i+1)\ell] = u \ \wedge \ y[i\ell : (i+1)\ell] = v\}\big|.$$

*Let $P_y(v) = \sum_u P_{x,y}(u,v)$ and, when $P_y(v) > 0$, let $P_{x|y}(u \mid v) = P_{x,y}(u,v)/P_y(v)$.*

These distributions induce a normalized disjoint-block conditional entropy rate.

▶ **Definition 12** (Disjoint $\ell$-block conditional entropy rate [17]). *With notation as in Definition 11, define*

$$H_\ell(x \mid y) = -\frac{1}{\ell} \sum_{v \in \Delta^\ell} P_y(v) \sum_{u \in \{0,1\}^\ell} P_{x|y}(u \mid v) \log P_{x|y}(u \mid v),$$

*with the convention that the inner term is 0 when $P_y(v) = 0$. For infinite sequences $X \in \{0,1\}^\infty$ and $Y \in \Delta^\infty$, define $H_\ell(X \mid Y) = \liminf_{k\to\infty} H_\ell\left(X \upharpoonright k\ell \mid Y \upharpoonright k\ell\right)$.*

Finite-state relative dimension is obtained by taking the infimum over block lengths.

▶ **Definition 13** (Finite-state relative dimension [17]). *For $X \in \{0,1\}^\infty$ and $Y \in \Delta^\infty$, define* $\dim_{\mathrm{FS}}^Y(X) = \inf_{\ell \geq 1} H_\ell(X \mid Y)$.

It is known that one may equivalently define $\dim_{\mathrm{FS}}^Y(X)$ using either disjoint-block or sliding-window conditional $\ell$-block entropy rates (with the same resulting value). [20]. Using relative dimension, we define relative independence (this is the notion used throughout the paper).

▶ **Definition 14** (Relative independence [17]). *Two sequences $X, Y \in \{0,1\}^\infty$ are* relatively independent *if $\dim_{\mathrm{FS}}^Y(X) = \dim_{\mathrm{FS}}(X)$ and $\dim_{\mathrm{FS}}^X(Y) = \dim_{\mathrm{FS}}(Y)$. More generally, for $k \geq 2$ and $(X_1, \ldots, X_k) \in (\{0,1\}^\infty)^k$, write $\widehat{X}_i = \langle X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k \rangle \in (\{0,1\}^{k-1})^\infty$, and say $(X_1, \ldots, X_k)$ is* mutually relatively independent *if for every $i$, $\dim_{\mathrm{FS}}^{\widehat{X}_i}(X_i) = \dim_{\mathrm{FS}}(X_i)$.*

Finally, we record the basic monotonicity that conditioning cannot increase finite-state dimension.

▶ **Lemma 15** ([17]). *For all $\ell$, all $X \in \{0,1\}^\infty$, and all oracles $Y$ over any finite alphabet,*

$H_\ell(X \mid Y) \leq H_\ell(X)$. *Consequently* $\dim_{\mathrm{FS}}^Y(X) \leq \dim_{\mathrm{FS}}(X)$.

### 2.4.1 A jointly normal source of bits

Several constructions below use a binary *source of bits* whose paired output is normal over $\{0,1\}^2$. Joint normality over the paired alphabet immediately yields maximal finite-state conditional block entropy in both directions.

▶ **Lemma 16.** *Let $W \in (\{0,1\}^2)^\infty$ be normal over the alphabet $\{0,1\}^2$, and let $A, B \in \{0,1\}^\infty$ be its coordinate projections, so $W = \langle A, B \rangle$. Then for every $\ell \geq 1$,*

$$H_\ell(A \mid B) = 1 \qquad \text{and} \qquad H_\ell(B \mid A) = 1.$$

*In particular,* $\dim_{\mathrm{FS}}^B(A) = \dim_{\mathrm{FS}}(A) = 1$ *and* $\dim_{\mathrm{FS}}^A(B) = \dim_{\mathrm{FS}}(B) = 1$.

## 3 Finite-state randomness extraction fails for a single source

In this section we prove the single-source non-extraction theorem. We construct a sequence $X$ of finite-state dimension $1/2$ and show that every finite-state transducer $T$ satisfies $\dim_{\mathrm{FS}}(T(X)) \leq 1/2$. The section proceeds in four steps: (i) we record basic technical lemmas (heavy-atom entropy and stability of normality under sparse padding), (ii) we prove

a structural SCC cycle-rate bound that controls the output produced by a transducer once its run is confined to a terminal strongly connected component, (iii) we give the stage-by-stage construction of $X$ and analyze it – first computing $\dim_{\mathrm{FS}}(X) = 1/2$, then using the SCC bound to show that each $T(X)$ has a heavy block occurring with asymptotic frequency at least $1/2$, forcing $\dim_{\mathrm{FS}}(T(X)) \leq 1/2$, and (iv) we show that the construction is efficient by giving a nearly linear-time algorithm to generate $X \upharpoonright N$.

## 3.1 Technical lemmas

We collect a few routine estimates that will be used repeatedly in the construction and analysis.

### 3.1.1 Heavy atom entropy and normality under sparse padding

We will repeatedly use the following simple estimate showing that a distribution on $\{0,1\}^\ell$ with a sufficiently heavy atom must have bounded (normalized) Shannon entropy.

▶ **Lemma 17.** *Fix $\ell \in \mathbb{N}$ and let $P$ be a distribution on $\Omega = \{0,1\}^\ell$. If $\max_{u\in\Omega} P(u) \geq \frac{1}{2} - \varepsilon$ for some $\varepsilon \in [0, \frac{1}{2}]$, then*

$$H(P) \;:=\; -\sum_{u\in\Omega} P(u)\log_2 P(u) \;\leq\; h\left(\tfrac{1}{2}-\varepsilon\right) \;+\; \left(\tfrac{1}{2}+\varepsilon\right)\ell.$$

The next lemma shows that normality is stable under inserting asymptotically smaller padding: if a normal sequence is partitioned into rapidly growing consecutive blocks and, after each block, we insert an auxiliary string whose length is asymptotically smaller than the corresponding block length, then the resulting interleaved sequence remains normal.

▶ **Lemma 18.** *Let $n_1 = 1$ and $n_i = (\sum_{j<i} n_j)^2$ for $i \geq 2$. Let $(c_i)_{i\geq 1}$ be positive integers with $c_i \leq i^{O(1)}$, and let $g : \mathbb{N} \to \mathbb{N}$ satisfy $g(i) = o(n_i)$. Let $R \in \{0,1\}^\infty$ be normal, and write $R = r_1 r_2 r_3 \cdots$ as consecutive chunks with $|r_i| = c_i n_i$. If $v_i \in \{0,1\}^*$ satisfies $|v_i| \leq g(i)$ for each $i$, then $R' = r_1 v_1 r_2 v_2 r_3 v_3 \cdots$ is normal.*

### 3.1.2 Structural core: cycle-rate bounds inside terminal SCCs

We isolate the key combinatorial mechanism that drives the construction. View the transition graph of an FST as a directed graph on $Q$, with one outgoing edge per input symbol from each state. Inside any strongly connected component (SCC), one can meaningfully talk about the *best possible average output rate* of a directed cycle. The crucial point is that this maximal cycle rate upper bounds the output produced along *any* walk that remains in the SCC, up to an additive constant depending only on the transducer. This is the graph-theoretic "rate control" phenomenon exploited by our diagonal construction.

▶ **Definition 19** (Maximal simple-cycle rate in an SCC)**.** *Let $T = (Q, \{0,1\}, \{0,1\}, q_0, \delta, \tau)$ and let $S$ be an SCC of the underlying directed graph of $T$. For a directed simple cycle $C$ in the labeled graph $E_T$ that stays in $S$, let $q_C$ be the $\leq$-least state on $C$. Let $\alpha_C \in \{0,1\}^{|C|}$ and $\beta_C \in \{0,1\}^*$ be the input/output labels of one traversal of $C$ starting at $q_C$. Define*

$$\mathrm{rate}(C) = \frac{|\beta_C|}{|C|}, \qquad \rho_S = \max\{\mathrm{rate}(C) : C \text{ a simple cycle in } S\}.$$

*Let $L = \max_{q\in Q,\ a\in\{0,1\}} |\tau(q,a)|$.*

Above, $\mathrm{rate}(C) \in [0, L]$. The quantity $\mathrm{rate}(C)$ is independent of the chosen starting vertex on $C$ (changing the start only cyclically rotates $\alpha$ and $\beta$). When we later speak of *choosing* a cycle $C$ in $S$ that attains $\rho_S$, we fix a canonical choice to break ties: among all directed simple cycles with $\mathrm{rate}(C) = \rho_S$, take one whose vertex-set has lexicographically least characteristic vector with respect to the fixed state order on $Q$. If this still leaves multiple cycles (e.g. distinct labeled cycles on the same vertex set), break ties by choosing the one with lexicographically least input label $\alpha_C$ among those, where $\alpha_C$ is read starting at $q_C$. We refer to this as *the* maximizing cycle in $S$ (this choice is used only for definiteness).

Our first structural lemma shows that, within a fixed strongly connected component (SCC), the maximal output rate over simple cycles bounds the output rate along any walk in that component. Apart from a short acyclic residue of a constant length, the walk decomposes into simple cycles whose average rates are all bounded by $\rho_S$.

▶ **Lemma 20** (Rate bound for arbitrary walks in an SCC). *Let $T = (Q, \{0,1\}, \{0,1\}, q_0, \delta, \tau)$ be a finite-state transducer. Fix a strongly connected component $S$ of the directed state graph of $T$, and let $\rho_S$ and $L$ be as in Definition 19. If $q, q' \in S$ and $w \in \{0,1\}^*$ labels a walk from $q$ to $q'$ that stays entirely inside $S$ (i.e. $q' = \widehat{\delta}(q, w)$ and every visited state lies in $S$), then*

$$|\widehat{\tau}(q, w)| \;\le\; \rho_S \cdot |w| \;+\; O_S(1),$$

*where $O_S(1)$ denotes a constant depending only on $|Q|$ and $L$ (in particular, independent of $w$).*

**Proof sketch.** Run $T$ on $w = a_1 \cdots a_t$ from $q$, obtaining the state walk $q_1 = q, \ldots, q_{t+1} = q'$. Decompose this walk into directed simple cycles plus a residual directed walk with no repeated vertices by repeatedly erasing the segment between two occurrences of the same state and recording the resulting simple cycle. The residual walk has length at most $|Q| - 1$.

Each recorded cycle lies in $S$, so by definition its average output per step is at most $\rho_S$, hence the total output over all cyclic parts is at most $\rho_S|w|$. The residual contributes at most $(|Q| - 1)L$ since it has at most $|Q| - 1$ steps and each step outputs at most $L$ bits. Therefore

$$|\widehat{\tau}(q, w)| \le \rho_S|w| + (|Q| - 1)L,$$

i.e. the claim holds with $O_S(1) = (|Q| - 1)L$. ◀

## 3.2 Main theorem and construction

We now prove the main theorem of this section: single-source randomness extraction fails for finite-state transducers, in the strong sense that there exists a sequence of finite-state dimension $\frac{1}{2}$ from which no transducer can produce output of dimension exceeding $\frac{1}{2}$.

▶ **Theorem 21** (Single-source non-extraction at dimension $\frac{1}{2}$). *There exists a binary sequence $X \in \{0,1\}^\infty$ such that*

$$\dim_{\mathrm{FS}}(X) = \frac{1}{2} \qquad \textit{and} \qquad \forall \textit{ binary FSTs } T, \quad \dim_{\mathrm{FS}}(T(X)) \le \frac{1}{2}.$$

### 3.2.1 Preliminary setup

We fix once and for all an effective enumeration $(T_e)_{e \ge 1}$ of all binary-output finite-state transducers with input alphabet $\{0,1\}$. We may assume (by enumerating machines in increasing number of states and then by tables) that the $e$-th transducer has at most $e$ states. We also fix a computable schedule $e(1), e(2), \ldots$ that visits every index infinitely often and satisfies $e(i) \le i$ (for example $1, 1, 2, 1, 2, 3, \ldots$).

We work on a rapidly growing scale. Define

$$n_1 = 1, \qquad n_i = \Big(\sum_{j<i} n_j\Big)^2 \quad (i \geq 2),$$

so $\sum_{j<i} n_j = \sqrt{n_i}$ and in particular $\sum_{j<i} n_j = o(n_i)$.

Finally, fix a normal sequence $R \in \{0,1\}^\infty$, which will serve as a source of "high-entropy" bits. At stage $i$ we will consume the next unused block of $R$ of a length prescribed by the construction.

The following simple estimate will be used repeatedly to justify that any polynomially bounded bookkeeping overhead (such as connector lengths or state-size parameters from the enumeration) is negligible compared to the rapidly growing stage scale $n_i$.

▶ **Lemma 22** (Polynomial overhead is negligible on the $n_i$ scale)**.** *Let $g : \mathbb{N} \to \mathbb{N}$ satisfy $g(i) \leq i^c$ for some constant $c$. Then*

$$\sum_{j<i} g(j)\, n_j \;=\; o(n_i) \qquad \text{and} \qquad \sum_{j<i} g(j) \;=\; o(n_i).$$

### 3.2.2 Construction of $X$

The construction is driven by the graph-theoretic structure of the transducer currently under attack. At the beginning of stage $i$, we first steer the transducer into a *terminal* strongly connected component (SCC), chosen canonically as follows: among all terminal SCCs reachable from the current state, let $B$ be the one whose set of states has lexicographically least characteristic vector with respect to the fixed state ordering $Q = \{q_1, \ldots, q_{|Q|}\}$ (i.e., we minimize the bitstring $(\mathbf{1}[q_1 \in B], \ldots, \mathbf{1}[q_{|Q|} \in B])$). We then work entirely inside this SCC. Inside $B$ we select a *simple* cycle of maximal average output rate and force the transducer to traverse it for a very long time. The SCC rate bound (Lemma 20) is the key combinatorial fact: it guarantees that, within a fixed SCC, the output produced on *any* walk that stays in the SCC is controlled (up to an additive constant) by the maximal cycle rate. This makes the predictable part of the construction dominate the $\ell$-block entropy of the output in a robust way: the domination holds uniformly as $n_i \to \infty$, with only an additive $O_T(1)$ slack depending on the fixed transducer $T$.

▶ **Definition 23** (Stage parameters and the sequence $X$)**.** *We define $X$ as*

$$X = r_1 p_1\, r_2 p_2\, r_3 p_3 \cdots,$$

*where stage $i$ consists of a random phase $r_i$ followed by a predictable phase $p_i$. Assume stages $< i$ have been defined and write*

$$X_{<i} = r_1 p_1 \cdots r_{i-1} p_{i-1}.$$

*Let $T = T_{e(i)} = (Q, \{0,1\}, \{0,1\}, q_0, \delta, \tau)$ and let $q^{\mathrm{cur}} = \widehat{\delta}(q_0, X_{<i})$ be the state of $T$ after reading $X_{<i}$.*

*Let $B$ be the canonically chosen* reachable terminal SCC *of $T$ from $q^{\mathrm{cur}}$, and let $v_i \in \{0,1\}^{\leq |Q|-1}$ be the input label of a fixed canonical shortest path[2] from $q^{\mathrm{cur}}$ to some state of $B$. Set*

$$q^{\mathrm{bot}} = \widehat{\delta}(q^{\mathrm{cur}}, v_i) \in B, \qquad \text{and let } S := B.$$

[2] We fix a deterministic tie-breaking for shortest paths: run BFS, processing states in increasing order and scanning outgoing edges in the order $0, 1$; the parent pointers from the first discovery define a unique shortest path. This is canonical and computable in $O(|Q|)$ time.

*Choose a directed simple cycle $C_i$ in $S$ maximizing the simple-cycle rate $\rho_S$ (Definition 19). Let $\alpha_i \in \{0,1\}^{m_i}$ and $\beta_i \in \{0,1\}^*$ be its input/output labels for one traversal from its base state. Define the predictable phase*

$$p_i := \alpha_i^{n_i}, \qquad \text{so} \qquad |p_i| = m_i n_i.$$

*Let $w_i \in \{0,1\}^{|p_i|}$ be the next $|p_i|$ unused bits of $R$. Since $S$ is terminal, $\widehat{\delta}(q^{\mathrm{bot}}, w_i) \in S$; let*

$$q' = \widehat{\delta}(q^{\mathrm{bot}}, w_i) \in S.$$

*Fix a canonical shortest path in $S$ from $q'$ to the base state of $C_i$; let $u_i \in \{0,1\}^{\leq |Q|-1}$ be its input label. Define the random phase*

$$r_i := v_i\, w_i\, u_i.$$

The words $v_i$ and $u_i$ are only used to enter the chosen terminal SCC and then place the machine at the chosen cycle base state; each has length at most $|Q|-1$ and is short compared to $|p_i|$. Indeed, since $|Q| \leq e(i) \leq i$ we have $|v_i|, |u_i| \leq |Q| - 1 \leq i$, and hence $|v_i|, |u_i| = o(n_i)$.

▶ **Lemma 24** (Lengths within a stage). *For every stage $i$,*

$$|r_i| = |p_i| + |v_i| + |u_i| \qquad \text{and} \qquad |v_i|, |u_i| \leq i.$$

*In particular, $|r_i| = |p_i| + o(n_i)$ and $|r_i|/|p_i| \to 1$.*

The above lemma follows from Definition 23 and $|Q| \leq e(i) \leq i$, since $|v_i|, |u_i| \leq |Q| - 1$.

### 3.2.3 Normality of the random phases

Let $W = r_1 r_2 r_3 \cdots$ be the infinite word obtained by concatenating the random-phase blocks from each stage. Intuitively, each $r_i$ consists mostly of a fresh chunk $w_i$ taken consecutively from the fixed normal source $R$, with only very short connector strings inserted to steer the transducer between SCC locations. Since these connectors have negligible length compared to the $w_i$ blocks, the overall concatenation $W$ should retain the block-frequency statistics of $R$ and hence remain normal.

▶ **Lemma 25** (Random-phase concatenation is normal). *The sequence $W$ is normal.*

### 3.2.4 Computing the dimension of $X$

The lower bound comes from mixing: asymptotically half of the disjoint blocks lie in the normal part $W$, so concavity forces the overall normalized block entropy to be at least $1/2$. The upper bound comes from a repeated-block phenomenon: at many stage boundaries, a single block occurs with disjoint frequency at least $1/2$, and Lemma 17 converts this into an entropy bound.

We begin by showing that, by construction, the random and predictable phases contribute asymptotically equal length to the prefixes of $X$.

▶ **Lemma 26** (Asymptotic balance of random and predictable mass). *Let $N_t = |r_1 p_1 \cdots r_t p_t|$. Then*

$$\frac{|r_1 \cdots r_t|}{N_t} \longrightarrow \frac{1}{2}, \qquad \text{and} \qquad \frac{|p_1 \cdots p_t|}{N_t} \longrightarrow \frac{1}{2}.$$

Now we prove the lower bound on $\dim_{\mathrm{FS}}(X)$.

▶ **Lemma 27** (Lower bound: $\dim_{\mathrm{FS}}(X) \geq \frac{1}{2}$)**.** *For every $\ell \in \mathbb{N}$, $H_\ell(X) \geq \frac{1}{2}$. Consequently, $\dim_{\mathrm{FS}}(X) \geq \frac{1}{2}$.*

Fix $\ell$. We work with the empirical $\ell$-block distribution obtained by *sliding-window* counts (i.e. counting each length-$\ell$ factor starting at every position); as remarked in the preliminaries this yields the same finite-state dimension as disjoint blocks.

For $N \geq \ell$, let $m(N) = N - \ell + 1$, and let $P_\ell^{(N)}$ denote the empirical distribution on $\{0,1\}^\ell$ obtained by counting all length-$\ell$ factors in the prefix $X \upharpoonright N$. Define the corresponding normalized entropy by

$$\overline{H}_\ell(X \upharpoonright N) = \frac{1}{\ell} H\Big(P_\ell^{(N)}\Big). \tag{1}$$

It is therefore enough to show that for every $\varepsilon > 0$ there exists $N_0$ such that for all $N \geq N_0$,

$$\overline{H}_\ell(X \upharpoonright N) \;\geq\; \frac{1}{2} - \varepsilon. \tag{2}$$

Fix $N \geq \ell$ and let $t = t(N)$ be the unique index such that $N_{t-1} < N \leq N_t$, where $N_s = |r_1 p_1 \cdots r_s p_s|$ and $N_0 = 0$. The prefix $X \upharpoonright N$ consists of the complete stages $1, \ldots, t-1$ followed by a (possibly empty) prefix of stage $t$.

We first show that we may discard a vanishing fraction of window starts whose length-$\ell$ windows are affected by stage bookkeeping (boundaries and short connector segments) without changing the empirical distribution by more than a quantity $\varepsilon_N$ in total variation, where $\varepsilon_N \to 0$ as $N \to \infty$.

▶ **Lemma 28.** *Let $\widetilde{P}_\ell^{(N)}$ be the empirical distribution obtained by restricting to window starts whose length-$\ell$ window lies entirely inside a $w_i$-segment or entirely inside a $p_i$-segment (i.e. we exclude windows that cross any segment boundary, and windows fully contained in connector segments $v_i$ or $u_i$). Then*

$$\big\|P_\ell^{(N)} - \widetilde{P}_\ell^{(N)}\big\|_{\mathrm{TV}} \;\longrightarrow\; 0 \qquad \text{as } N \to \infty.$$

Next we identify $\widetilde{P}_\ell^{(N)}$ as a mixture of the window distribution inside the random segments and the window distribution inside the predictable segments. The key points are that (a) the mixture weight on the random part is at least $\frac{1}{2} - \varepsilon_N$ for some $\varepsilon_N \to 0$, and (b) the random-part window distribution is asymptotically uniform because the $w$-bits form a normal subsequence of the source.

▶ **Lemma 29.** *Let $P_{\ell,\mathrm{rand}}^{(N)}$ be the empirical distribution of length-$\ell$ windows whose starting positions are such that the window lies entirely inside some $w_i$-segment, and let $P_{\ell,\mathrm{pred}}^{(N)}$ be the analogous distribution for windows lying entirely inside some $p_i$-segment. Let $\theta_N$ be the fraction of all window starts counted by $\widetilde{P}_\ell^{(N)}$ whose corresponding window lies entirely inside a $w_i$-segment. Then*

$$\widetilde{P}_\ell^{(N)} \;=\; \theta_N\, P_{\ell,\mathrm{rand}}^{(N)} \;+\; (1-\theta_N)\, P_{\ell,\mathrm{pred}}^{(N)}. \tag{3}$$

*Moreover, there exists a function $\varepsilon_N \geq 0$ with $\varepsilon_N \to 0$ such that $\theta_N \geq \frac{1}{2} - \varepsilon_N$ as $N \to \infty$, and $H(P_{\ell,\mathrm{rand}}^{(N)}) \to \ell$ as $N \to \infty$.*

We now complete the proof of Lemma 27.

**Proof.** We now combine the mixture lower bound with concavity of Shannon entropy and then transfer the resulting bound from $\widetilde{P}_\ell^{(N)}$ back to $P_\ell^{(N)}$ using a standard continuity estimate.

Concavity of Shannon entropy applied to (3) yields

$$H\Big(\widetilde{P}_\ell^{(N)}\Big) \;\geq\; \theta_N\, H\Big(P_{\ell,\mathrm{rand}}^{(N)}\Big) + (1-\theta_N)\, H\Big(P_{\ell,\mathrm{pred}}^{(N)}\Big) \;\geq\; \theta_N\, H\Big(P_{\ell,\mathrm{rand}}^{(N)}\Big), \tag{4}$$

using $H(P_{\ell,\mathrm{pred}}^{(N)}) \geq 0$. Using Lemma 29, this implies that there exists a function $\varepsilon_N'' \geq 0$ with $\varepsilon_N'' \to 0$ such that $H(\widetilde{P}_\ell^{(N)}) \geq \frac{\ell}{2} - \varepsilon_N''\,\ell$.

We now transfer the entropy lower bound from $\widetilde{P}_\ell^{(N)}$ to $P_\ell^{(N)}$ using a standard continuity bound for Shannon entropy on a finite alphabet. Let $\delta_N = \|P_\ell^{(N)} - \widetilde{P}_\ell^{(N)}\|_{\mathrm{TV}}$ and note that the alphabet size is $M = |\{0,1\}^\ell| = 2^\ell$. We use the following inequality from [18]: for any two distributions $P, Q$ on an alphabet of size $M$ and any $\delta \in [0, 1/2]$ with $\|P - Q\|_{\mathrm{TV}} \leq \delta$,

$$|H(P) - H(Q)| \;\leq\; h\,(\delta) + \delta \log_2(M-1) \;\leq\; h\,(\delta) + \delta\,\ell, \tag{5}$$

where $h\,(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function. Applying (5) with $P = P_\ell^{(N)}$, $Q = \widetilde{P}_\ell^{(N)}$, and $\delta = \delta_N$, and using $\delta_N \to 0$ from Lemma 28, we obtain

$$H\Big(P_\ell^{(N)}\Big) \;\geq\; H\Big(\widetilde{P}_\ell^{(N)}\Big) - \big(h\,(\delta_N) + \delta_N \ell\big) \;=\; H\Big(\widetilde{P}_\ell^{(N)}\Big) - \varepsilon_N'''\,\ell, \tag{6}$$

where $\varepsilon_N''' := \frac{h\,(\delta_N)}{\ell} + \delta_N$ satisfies $\varepsilon_N''' \to 0$ as $N \to \infty$.

Combining the preceding bounds yields: for every $\varepsilon > 0$ there exists $N_0$ such that for all $N \geq N_0$,

$$H\Big(P_\ell^{(N)}\Big) \;\geq\; \Big(\frac{1}{2} - \varepsilon\Big)(\ell - \varepsilon\ell) \;-\; \varepsilon\ell \;\geq\; \Big(\frac{1}{2} - 3\varepsilon\Big)\ell.$$

Dividing by $\ell$ and using (1) gives (2). Hence $\liminf_{N\to\infty} \overline{H}_\ell(X \upharpoonright N) \geq \frac{1}{2}$, and therefore $H_\ell(X) \geq \frac{1}{2}$. ◀

The next lemma gives the matching upper bound on $\dim_{\mathrm{FS}}(X)$ by exploiting the long periodic predictable phases, which force a heavy $\ell$-block and hence low block entropy along infinitely many stage endpoints.

▶ **Lemma 30** (Upper bound: $\dim_{\mathrm{FS}}(X) \leq \frac{1}{2}$)**.** *For every $\varepsilon > 0$ there exists $\ell$ such that $H_\ell(X) \leq \frac{1}{2} + \varepsilon$. Consequently,* $\dim_{\mathrm{FS}}(X) \leq \frac{1}{2}$.

**Proof.** Fix a particular transducer index $e_0$ (for concreteness $e_0 = 1$). The schedule visits $e_0$ infinitely often. For stages $i$ with $e(i) = e_0$, the chosen cycle input lengths $m_i = |\alpha_i|$ are bounded by $|Q_{e_0}|$, so some value $m^*$ occurs infinitely often along those stages.

Fix $j \in \mathbb{N}$ and set $\ell = jm^*$. Consider infinitely many stages $i$ with $e(i) = e_0$ and $|\alpha_i| = m^*$. For such a stage, $p_i = \alpha_i^{n_i}$ is a repetition of a word of length $m^*$, and hence every disjoint $\ell$-block fully contained in $p_i$ equals the same word $u_{i,j} \in \{0,1\}^\ell$. The number of such disjoint $\ell$-blocks inside $p_i$ is at least $\lfloor n_i/j \rfloor - 2$.

Let $N_i = |r_1 p_1 \cdots r_i p_i|$. By the stage length relation (Lemma 24) and polynomial-negligibility (Lemma 22), the total connector mass up to stage $i$ is $o(|p_i|)$, and the total mass of all earlier stages is also $o(|p_i|)$. Consequently, $N_i \;\leq\; 2|p_i| + o(|p_i|)$. Therefore the total number of disjoint $\ell$-blocks in $X \upharpoonright N_i$ is at most

$$\left\lfloor \frac{N_i}{\ell} \right\rfloor \;\leq\; \frac{2|p_i| + o(|p_i|)}{jm^*} \;=\; \frac{2n_i}{j} + o\Big(\frac{n_i}{j}\Big).$$

Consequently, along this infinite subsequence of stages,

$$P(u_{i,j}, X \restriction N_i) \;\geq\; \frac{\lfloor n_i/j \rfloor - 2}{(2n_i/j) + o(n_i/j)} \;\longrightarrow\; \frac{1}{2}$$

Hence there is a sequence $\varepsilon_i \to 0$ such that, along this infinite subsequence,

$$P(u_{i,j}, X \restriction N_i) \geq \tfrac{1}{2} - \varepsilon_i.$$

Applying Lemma 17 gives

$$H_\ell(X \restriction N_i) \leq \frac{1}{2} + \varepsilon_i + \frac{1}{\ell}.$$

The definition of $H_\ell(X)$ uses prefixes of length $k\ell$; replacing $N_i$ by the nearest multiple of $\ell$ changes at most one disjoint $\ell$-block and does not affect the liminf. Hence $H_\ell(X) \leq \frac{1}{2} + \frac{1}{\ell}$. Choosing $j$ large enough so that $1/\ell < \varepsilon$ gives $H_\ell(X) \leq \frac{1}{2} + \varepsilon$, completing the proof. ◀

Combining the preceding bounds, we obtain the exact value of the finite-state dimension of $X$.

▶ **Corollary 31.** $\dim_{\mathrm{FS}}(X) = \frac{1}{2}$.

**Proof.** Combine Lemmas 27 and 30. ◀

### 3.2.5 Bounding every transducer output via SCC cycle rates

We now show that no transducer can raise the finite-state dimension above $\frac{1}{2}$ on the constructed input. This is where the SCC combinatorics is used most transparently: in the stage devoted to a given transducer, we force the machine to spend a very long time traversing a cycle of maximal average output rate in its current SCC. The SCC rate lemma then implies that the output produced on the preceding SCC-internal walk is *at most* the output produced on this maximal-rate cycle, up to an additive constant depending only on the machine. As a result, at the end of such a stage, approximately half of the output consists of one repeated block, forcing low block entropy.

The next lemma is the point where the SCC cycle-rate analysis enters the diagonalization: in the stage devoted to a fixed transducer, the output produced on the SCC-internal part of the "random" phase cannot exceed (up to an additive constant) the output produced on the forced traversal of a maximal-rate cycle.

▶ **Lemma 32.** *Fix $e$ and consider a stage $i$ with $e(i) = e$. Let $T = T_e$, with state set $Q$ and output function $\tau$. Let $b_i$ be the output produced by $T$ while reading $p_i$ at stage $i$. Let $S$ be the SCC selected at stage $i$, and let $C_i$ be the chosen maximal-rate simple cycle in $S$ with input/output labels $(\alpha_i, \beta_i)$. Write the random phase as $r_i = y_i x_i$, where $x_i$ is the SCC-internal suffix of $r_i$ that stays inside $S$ and ends at the base state of $C_i$. Let $s_i$ be the output produced while reading $x_i$. Then $b_i = \beta_i^{n_i}$ and $|s_i| \leq |b_i| + O_T(1)$, where $O_T(1)$ depends only on $T$.*

Using the superlinear growth of the stage lengths, we next compare the forced cycle output $b_i$ to the entire output prefix at the end of the stage, and show that $b_i$ occupies asymptotically half of it.

▶ **Lemma 33.** *Fix $e$ and let $T = T_e$. Let $N_i^{\mathrm{out}}$ be the output length of $T(X)$ at the end of a stage $i$ with $e(i) = e$. If $T(X)$ is infinite, then, along all sufficiently late such stages, $|b_i|/N_i^{\mathrm{out}} \to \frac{1}{2}$. If the eventual maximizing cycle has empty output, then $T(X)$ is finite.*

We can now turn the half-output periodicity into an entropy bound, concluding that the output of any fixed transducer has finite-state dimension at most $\frac{1}{2}$.

▶ **Lemma 34.** *For every binary FST $T$, $\dim_{\mathrm{FS}}(T(X)) \le \frac{1}{2}$.*

**Proof.** Fix $T = T_e$. If $T(X)$ is finite, then $\dim_{\mathrm{FS}}(T(X)) = 0$. Assume $T(X)$ is infinite.

After the first stage devoted to $T$, the run of $T$ has entered a terminal SCC. Since this SCC is terminal, the run never leaves it. Hence, for all sufficiently late stages $i$ with $e(i) = e$, the selected SCC is the same, and by the canonical tie-breaking rule the chosen maximizing cycle is also the same. Let $\beta^*$ be its output label. If $\beta^* = \lambda$, then the maximal cycle rate in this terminal SCC is 0. Since the SCC is strongly connected, every edge in it must then have empty output, and $T(X)$ would be finite. Thus $\beta^* \neq \lambda$.

Fix $j \in \mathbb{N}$ and set $\ell = j|\beta^*|$. For all sufficiently late stages $i$ with $e(i) = e$, we have $b_i = (\beta^*)^{n_i}$. Inside $b_i$, every disjoint $\ell$-block fully contained in $b_i$ equals the same word $v_{i,j} \in \{0,1\}^\ell$. By Lemma 33, the ratio $|b_i|/N_i^{\mathrm{out}}$ tends to $\frac{1}{2}$ along these stages. The same disjoint-block counting used in Lemma 30 therefore gives

$$P\big(v_{i,j}, T(X) \restriction N_i^{\mathrm{out}}\big) \;\ge\; \frac{1}{2} - \varepsilon_i,$$

for some $\varepsilon_i = \varepsilon_i(j) \ge 0$ with $\varepsilon_i \to 0$ along this infinite subsequence of stages. Applying Lemma 17 gives

$$H_\ell\big(T(X) \restriction N_i^{\mathrm{out}}\big) \;\le\; \frac{1}{2} + \varepsilon_i + \frac{1}{\ell}.$$

The definition of $H_\ell(T(X))$ uses prefixes whose lengths are multiples of $\ell$; replacing $N_i^{\mathrm{out}}$ by the nearest multiple of $\ell$ changes at most one disjoint $\ell$-block and does not affect the liminf. Hence $H_\ell(T(X)) \le \frac{1}{2} + \frac{1}{\ell}$. Letting $j \to \infty$, so that $\ell = j|\beta^*| \to \infty$, yields $\dim_{\mathrm{FS}}(T(X)) \le \frac{1}{2}$. ◀

We now assemble the construction and the preceding dimension bounds to complete the proof of the main theorem.

**Proof of Theorem 21.** Definition 23 defines $X$. Corollary 31 gives $\dim_{\mathrm{FS}}(X) = \frac{1}{2}$. Lemma 34 gives $\dim_{\mathrm{FS}}(T(X)) \le \frac{1}{2}$ for every binary FST $T$. ◀

We show that the same diagonal cycle-guided construction can be tuned to any prescribed rational finite-state dimension.

▶ **Theorem 35** (Single-source non-extraction at arbitrary rational dimension)**.** *Let $s \in (0,1)$ be rational. There exists a binary sequence $X \in \{0,1\}^\infty$ such that*

$$\dim_{\mathrm{FS}}(X) = s \qquad \text{and} \qquad \forall \text{ binary FSTs } T, \;\; \dim_{\mathrm{FS}}(T(X)) \le s.$$

The construction is identical to the $\frac{1}{2}$-case, except that each stage is rebalanced so that the random part contributes an asymptotic $s$-fraction of the input and the predictable part contributes an asymptotic $(1-s)$-fraction. Write $s = \frac{a}{a+b}$ with integers $a, b \ge 1$. At stage $i$ for the scheduled transducer $T_{e(i)}$, from the current state we steer into a canonically chosen reachable terminal SCC $S$, select in $S$ a directed simple cycle $C_i$ of maximal simple-cycle output rate with input/output labels $(\alpha_i, \beta_i)$ (with $|\alpha_i| = m_i$), and then concatenate: (i) an SCC-internal random chunk $w_i$ consisting of the next $am_in_i$ unused symbols of a fixed normal source, followed by a shortest SCC-internal connector returning to the base of $C_i$, and (ii) a predictable block $p_i := \alpha_i^{bn_i}$. Thus each stage contributes $|w_i| = am_in_i$ symbols

from the normal source and $|p_i| = bm_i n_i$ symbols of forced periodic input (up to $o(n_i)$ steering/connector overhead), and the final sequence is $X = r_1 p_1 r_2 p_2 \cdots$ with $r_i := w_i u_i$ (absorbing any additional $o(n_i)$ bookkeeping). The rest of the proof follows the argument for Theorem 21 with the role of $\frac{1}{2}$ replaced by $s$ throughout.

## 3.3 Nearly linear-time computation of the constructed sequence

We show that the diagonal sequence $X$ from Definition 23 is effectively computable with nearly linear overhead: on input $N$ one can output the prefix $X \upharpoonright N$ in time $O\left(N \cdot (\log N)^c\right)$. The key point is that the stage lengths grow so quickly that any rescanning work done once per stage sums to $O(N)$.

### Model of computation and "nearly linear time"

We measure time on a (log-cost) RAM model, as is standard in explicit normality constructions (see, e.g., [14]). The cost of basic arithmetic and memory operations is proportional to the bit-length of the operands/addresses involved. In particular, manipulating integers of magnitude at most $N$ costs $O(\log N)$ per operation, and writing $N$ output bits necessarily costs $\Omega(N)$ time. We say that an algorithm runs in *nearly linear time* if it runs in time $O\left(N \cdot (\log N)^c\right)$ for some fixed constant $c$.

### An explicit normal source

For the normal sequence $R$ used in the construction, we may take the binary Champernowne sequence $C_2 = 0\,1\,10\,11\,100\,101\,110\,111\,1000\cdots$, obtained by concatenating the binary representations of $1, 2, 3, \ldots$ (without leading zeros). It is well known that $C_2$ is normal in base 2 [4]. The following lemma records that our chosen explicit normal source, the binary Champernowne sequence $C_2$, admits prefix generation with essentially optimal overhead (up to a $\log N$ factor) in the log-cost RAM model.

▶ **Lemma 36** (Computing the binary Champernowne prefix in nearly linear time)**.** *There is an algorithm that, on input $N$, outputs the first $N$ bits of $C_2$ in time $O\big(N \log N\big)$ in the log-cost RAM model.*

### A convenient enumeration of transducers

Fix an encoding of a binary-input, binary-output transducer $T = (Q, \{0,1\}, \{0,1\}, q_0, \delta, \tau)$ by listing: (i) the number of states $|Q|$, (ii) the start state $q_0$, (iii) the transition table $\delta(q,a)$ for all $q \in Q$ and $a \in \{0,1\}$, and (iv) the output table $\tau(q,a)$ for all $q, a$, encoded in any self-delimiting format. This yields an effective enumeration $(T_e)_{e \geq 1}$.

▶ **Lemma 37** (Efficient access to $T_e$)**.** *There is an effective enumeration $(T_e)_{e \geq 1}$ of all binary-output FSTs such that, on input $e$, one can decode a full description of $T_e$ (in particular its tables $(\delta, \tau)$) in time polynomial in $e$ in the log-cost RAM model. Moreover, we may assume that $T_e$ has at most $e$ states.*

### A computable diagonal schedule

We also fix the schedule $e(1), e(2), \ldots$ used in the diagonalization by listing indices along successive diagonals: for each $m \geq 1$, output the block $1, 2, \ldots, m$ and then move to $m+1$. Equivalently, $e(i)$ is defined by the unique $m \geq 1$ such that $\binom{m}{2} < i \leq \binom{m+1}{2}$ and then $e(i) = i - \binom{m}{2}$. This schedule visits every index infinitely often and is easily verified to be computed in time polynomial in $\log i$.

▶ **Theorem 38** (Nearly linear-time constructability). *Let $X$ be the sequence defined by Definition 23, and take $R = C_2$ (or any fixed normal sequence whose prefixes are computable in nearly linear time). There is an algorithm that, on input $N$, outputs $X \upharpoonright N$ in time $O\big(N \cdot (\log N)^c\big)$ for a fixed constant $c$ in the log-cost RAM model.*

**Proof sketch.** On input $N$, generate $X$ stage-by-stage until $N$ bits are written, and then truncate. At stage $i$, compute $e = e(i)$ and decode $T_{e(i)} = (Q, \{0,1\}, \{0,1\}, q_0, \delta, \tau)$ (Lemma 37). Compute the current state $q^{\mathrm{cur}} = \widehat{\delta}(q_0, X_{<i})$ by a single left-to-right scan of the already written prefix $X_{<i}$, updating the state using $\delta$. Compute the SCC decomposition of the directed state graph of $T_{e(i)}$ and its condensation DAG (e.g. Tarjan's linear-time SCC algorithm [21]). Among all *terminal* SCCs reachable from the SCC of $q^{\mathrm{cur}}$, choose $B$ canonically by minimizing the characteristic vector $(\mathbf{1}[q_1 \in B], \ldots, \mathbf{1}[q_{|Q|} \in B])$ in lexicographic order with respect to the fixed state ordering $Q = \{q_1, \ldots, q_{|Q|}\}$. Compute a shortest path from $q^{\mathrm{cur}}$ into the chosen SCC $B$ and output its input label $v_i$, landing at $q^{\mathrm{bot}} \in B$. Inside $B$, compute a maximum mean-weight cycle with edge weights $|\tau(q,a)|$ (equivalently, a cycle of maximal average output rate) using Karp's algorithm [10], extract a maximizing *simple* cycle $C_i$ with input/output labels $(\alpha_i, \beta_i)$, and set $p_i = \alpha_i^{n_i}$. Stream the next $|p_i|$ fresh bits from the normal source $R = C_2$ as $w_i$ while simultaneously simulating $T_{e(i)}$ from $q^{\mathrm{bot}}$ to find the state $q' = \widehat{\delta}(q^{\mathrm{bot}}, w_i) \in B$. Then compute and output a shortest path within $B$ from $q'$ to the base state of $C_i$, with input label $u_i$. Finally output the predictable block $p_i$. Thus each stage outputs $r_i = v_i w_i u_i$ followed by $p_i = \alpha_i^{n_i}$.

Writing $N$ bits costs $O(N)$ time, and streaming $R = C_2$ costs $O(N \log N)$ by Lemma 36. Per-stage graph work (SCC decomposition, reachability in the condensation DAG, canonical choice of $B$, BFS for $v_i, u_i$, and maximum-mean-cycle computation inside $B$) is polynomial in $|Q_{e(i)}|$. Since $|p_i| \geq n_i$ and $n_{i+1} \geq n_i^2$, the number of completed stages before reaching length $N$ is $t = O(\log \log N)$, so the total graph overhead over all stages is polylogarithmic in $N$ (absorbed by $N(\log N)^c$). The remaining overhead is rescanning $X_{<i}$ once per stage to compute $q^{\mathrm{cur}}$. Fast growth gives $\sum_{i \leq t} |X_{<i}| = O(N)$, so total rescan time is $O(N(\log N)^c)$ in the log-cost RAM model. ◀

## 4 Randomness extraction fails for multi-input transducers

This section establishes the multi-source analogue of the single-source non-extraction result. We do this by treating two independence regimes for the input streams. In Section 4.1 we assume the sources are independent in the sense of Definition 7, and show that no multi-input finite-state transducer can raise their finite-state dimension. In the final subsection we consider a second, finite-state notion of independence, namely finite-state relative independence (Definition 14), and prove that the same impossibility of extraction persists.

### 4.1 Independent sources defeating all multi-input transducers

In this section we construct two sources $X, Y$ that are independent in the sense of Definition 7, satisfy $\dim_{\mathrm{FS}}(X) = \dim_{\mathrm{FS}}(Y) = \frac{1}{2}$, and nevertheless defeat every two-input finite-state transducer in the sense that no such transducer can output a sequence of finite-state dimension greater than $\frac{1}{2}$ from $(X, Y)$.

Recall the fast-growing sequence $(n_i)$ from Section 3. In the multi-source arguments we repeatedly need the same kind of estimate: all earlier stages and all per-stage bookkeeping costs (connectors, boundary effects in block counts, and the short "stage description" information used later to recover the stage parsing) are negligible compared to the current stage. We will use without further comment that $n_{i+1} \geq n_i^2$, so $n_i$ grows super-exponentially and

the number of completed stages below a prefix of length $N$ is $O(\log\log N)$. We will also use Lemma 22, which says that any polynomially bounded overhead from earlier stages is $o(n_i)$ on the $i$-th stage scale.

### 4.1.1 Main theorem

We now state the main result of this subsection.

▶ **Theorem 39** (Two-source independent non-extraction at $\frac{1}{2}$)**.** *There exist sequences $X, Y \in \{0,1\}^\infty$ such that*

$$\dim_{\mathrm{FS}}(X) = \dim_{\mathrm{FS}}(Y) = \frac{1}{2}, \qquad X, Y \text{ are independent},$$

*and for every two-input FST $T$, $\dim_{\mathrm{FS}}(T(X,Y)) \le \frac{1}{2}$.*

**Proof.** Let $A, B \in \{0,1\}^\infty$ be normal and independent as in Lemma 10 (Definition 7). Fix an effective enumeration $(T_e)_{e\ge 1}$ of all two-input binary-output FSTs with input alphabet $\{0,1\}^2$, and a computable schedule $e(1), e(2), \dots$ visiting each $e$ infinitely often with $e(i) \le i$. Use the same fast-growing scale $(n_i)$ as in the single-source construction.

**Construction.** We build $Z \in (\{0,1\}^2)^\infty$ in stages $Z = r_1 p_1\, r_2 p_2 \cdots$ exactly as in Definition 23, with the only change that the source bits now come from the *paired* normal source $\langle A, B\rangle$: at stage $i$, for $T = T_{e(i)}$ we steer from the current state into the canonically chosen reachable terminal SCC $S$ via a shortest word $v_i \in (\{0,1\}^2)^{\le |Q|-1}$, choose in $S$ a maximal-rate simple cycle $C_i$ with labels $(\alpha_i, \beta_i)$ and $|\alpha_i| = m_i$, set $p_i := \alpha_i^{n_i}$ (so $|p_i| = m_i n_i =: L_i$), take the next $L_i$ unused bits $u_i \in \{0,1\}^{L_i}$ from $A$ and $v_i' \in \{0,1\}^{L_i}$ from $B$, let $w_i := \langle u_i, v_i'\rangle \in (\{0,1\}^2)^{L_i}$, and finally take the canonical shortest connector $c_i \in (\{0,1\}^2)^{\le |Q|-1}$ inside $S$ from the post-$w_i$ state to the base of $C_i$; define $r_i := v_i w_i c_i$. Now define $X, Y \in \{0,1\}^\infty$ by coordinate projection: $(X(t), Y(t)) = Z(t)$ for all $t$.

**Finite-state dimensions of $X$ and $Y$.** The verification is identical to the single-source argument (Lemmas 25, 27, 30), applied to each coordinate. Indeed, $X$ has the stage form $X = r_1^{(1)} p_1^{(1)} r_2^{(1)} p_2^{(1)} \cdots$, where $p_i^{(1)} = (\alpha_i^{(1)})^{n_i}$ and $r_i^{(1)} = v_i^{(1)} u_i c_i^{(1)}$. The concatenation $u_1 u_2 \cdots$ is a contiguous subsequence of $A$, hence normal, and the inserted padding between consecutive $u_i$ blocks has length $O(|Q_{e(i)}|) = i^{O(1)} = o(n_i)$; thus Lemma 18 gives that $r_1^{(1)} r_2^{(1)} \cdots$ is normal, exactly as in Lemma 25. With this normality input, the sliding-window mixture lower bound and the heavy-atom upper bound carry over verbatim (the same negligible-boundary/connector estimates, using the same fast growth of $(n_i)$), yielding $\dim_{\mathrm{FS}}(X) = \frac{1}{2}$. The same argument gives $\dim_{\mathrm{FS}}(Y) = \frac{1}{2}$.

**Bounding $\dim_{\mathrm{FS}}(T(X,Y))$.** Fix any two-input FST $T = T_e$. Consider stages $i$ with $e(i) = e$. In such a stage, by construction $T$ begins reading $p_i = \alpha_i^{n_i}$ at the base of the chosen maximal-rate cycle, so the output produced on $p_i$ equals $b_i := \beta_i^{n_i}$. Let $s_i$ be the output produced while reading the SCC-internal suffix $w_i c_i$ of the random phase. The SCC rate bound (Lemma 20) applied inside the chosen terminal SCC $S$ gives

$$|s_i| \;\le\; \rho_S(|w_i| + |c_i|) + O_T(1) \;=\; \rho_S |p_i| + o(|p_i|) + O_T(1) \;=\; |b_i| + o(|b_i|),$$

exactly as in Lemma 32. As in Lemma 33, output produced before stage $i$ is $o(|b_i|)$ by the superlinear stage growth, and the steering prefix $v_i$ contributes only $o(|b_i|)$ output. Hence,

along infinitely many such stages with $|b_i| > 0$, the block $b_i$ occupies asymptotically at least half of the output prefix at the end of the stage. After the first stage devoted to $T$, the run of $T$ has entered a terminal SCC, and it never leaves it. Hence, for all sufficiently late stages $i$ with $e(i) = e$, the selected SCC is the same, and by the canonical tie-breaking rule the chosen maximizing cycle is also the same. Let $\beta^*$ be its output label. If $\beta^* = \lambda$, then the maximal cycle rate in this terminal SCC is 0, so every edge in the SCC has empty output, and $T(X, Y)$ is finite after a bounded prefix. Otherwise, for each fixed $j$ set $\ell = j|\beta^*|$. The same disjoint-block heavy-atom argument from Lemma 34, applied inside $b_i = (\beta^*)^{n_i}$, gives $H_\ell(T(X, Y)) \le \frac{1}{2} + 1/\ell$. Letting $j \to \infty$ yields $\dim_{\mathrm{FS}}(T(X, Y)) \le \frac{1}{2}$.

**Independence.** Fix $n, m$, and let $t = t(n, m)$ be the largest stage that intersects either $X \upharpoonright n$ or $Y \upharpoonright m$. By the same stage-growth estimate used throughout the single-source proof, $t = O(\log\log(n + m))$.

Define a short advice string $D_{n,m}$ as follows. For every stage $i \le t$, encode the state $q_i^{\mathrm{cur}}$ of $T_{e(i)}$ at the beginning of stage $i$, and the state $q_i'$ reached after reading the source block $w_i$ inside the chosen SCC. Since $|Q_{e(i)}| \le i^{O(1)}$ and $e(i) \le i$, this costs $O(\log i)$ bits per stage. Therefore $|D_{n,m}| = O(t \log t) = O(\log n + \log m)$.

Given $D_{n,m}$, the fixed enumeration, and the canonical tie-breaking rules, all non-source parts of the construction up to stage $t$ are computable. Indeed, from $q_i^{\mathrm{cur}}$ one recomputes the chosen reachable terminal SCC, the steering word into it, and the maximal-rate cycle used for the predictable block. From $q_i'$ one recomputes the canonical connector back to the base of that cycle. Hence, up to the two prefix cut positions $n, m$, the only incomputable information in $X$ comes from the corresponding prefix of $A$, and the only incomputable information in $Y$ comes from the corresponding prefix of $B$.

Let $a = a(n, D_{n,m})$ be the number of bits of $A$ used in the source chunks appearing in $X \upharpoonright n$, and let $b = b(m, D_{n,m})$ be the analogous number for $B$ and $Y \upharpoonright m$. Then $a \le n$ and $b \le m$, and the preceding reconstruction gives $K(X \upharpoonright n) \le K(A \upharpoonright a) + O(\log n + \log m)$, and $K(Y \upharpoonright m) \le K(B \upharpoonright b) + O(\log n + \log m)$. Conversely, from $X \upharpoonright n$, $Y \upharpoonright m$, and $D_{n,m}$ one can delete the computable steering, connector, and predictable portions and recover $A \upharpoonright a$ and $B \upharpoonright b$. Thus $K(A \upharpoonright a,\ B \upharpoonright b) \le K(X \upharpoonright n,\ Y \upharpoonright m) + O(\log n + \log m)$. Using the independence of $A$ and $B$, we obtain

$$\begin{aligned} K(X \upharpoonright n,\ Y \upharpoonright m) &\ge K(A \upharpoonright a,\ B \upharpoonright b) - O(\log n + \log m) \\ &\ge K(A \upharpoonright a) + K(B \upharpoonright b) - O(\log n + \log m) \\ &\ge K(X \upharpoonright n) + K(Y \upharpoonright m) - O(\log n + \log m). \end{aligned}$$

Therefore $X$ and $Y$ are independent in the sense of Definition 7. ◀

We remark that the same construction extends to any fixed $k \ge 2$ and any prescribed rational target dimension.

▶ **Theorem 40** ($k$-source independent non-extraction at rational dimension $s$)**.** *Fix $k \ge 2$ and let $s \in (0, 1)$ be rational. There exist sequences $X_1, \dots, X_k \in \{0,1\}^\infty$ such that $(X_1, \dots, X_k)$ is independent, each $X_j$ has $\dim_{\mathrm{FS}}(X_j) = s$, and for every $k$-input finite-state transducer $T$,*

$$\dim_{\mathrm{FS}}(T(X_1, \dots, X_k)) \le s.$$

To construct the sequences, replace the paired alphabet $\{0,1\}^2$ by $\{0,1\}^k$ and fix $k$ normal sources $A_1, \dots, A_k$ that are $k$-way independent. Write $s = \frac{a}{a+b}$ with $a, b \in \mathbb{N}$. The construction and analysis then follow by rebalancing the random and predictable portions of each stage in the ratio $a : b$ (as in the single-source rational-$s$ case) and by carrying out the same argument over the product alphabet $\{0,1\}^k$, with only routine $k$-ary modifications.

## 4.2 Finite-state relatively independent sources defeating all multi-input transducers

In Section 4.1 we constructed independent sources from which no finite-state transducer can extract higher finite-state dimension. In this final section we record a stronger statement: the impossibility persists even when the sources are *relatively independent* in the sense of Definition 14, i.e., each source retains its finite-state dimension even when the other source(s) are available as an oracle in the finite-state sense.

### 4.2.1 Two relatively independent sources defeat all two-input transducers

We first treat the case of two sources; the $k$-source extension follows after.

▶ **Theorem 41** (Two relatively independent sources of dimension $\frac{1}{2}$ defeat all two-input transducers). *There exist sequences $X, Y \in \{0,1\}^\infty$ such that*

$$\dim_{\mathrm{FS}}(X) = \dim^{Y}_{\mathrm{FS}}(X) = \frac{1}{2}, \qquad \dim_{\mathrm{FS}}(Y) = \dim^{X}_{\mathrm{FS}}(Y) = \frac{1}{2}.$$

*and for every two-input binary-output finite-state transducer $T$, $\dim_{\mathrm{FS}}(T(X,Y)) \le \frac{1}{2}$.*

**Proof.** Fix a source $W \in (\{0,1\}^2)^\infty$ normal over $\{0,1\}^2$, and write $W = \langle A, B\rangle$ for its coordinate projections. Fix an effective enumeration $(T_e)_{e\ge 1}$ of two-input binary-output finite-state transducers $T_e = (Q, \{0,1\}^2, \{0,1\}, q_0, \delta, \tau)$ and a computable schedule $e(1), e(2), \dots$ visiting every $e$ infinitely often with $e(i) \le i$. Use the same fast-growing scale $(n_i)$ as in Sections 3–4.1.

**Construction.** Construct a paired stream $Z \in (\{0,1\}^2)^\infty$ by applying the single-source stage construction (Definition 23) verbatim over input alphabet $\{0,1\}^2$, with the normal source $R$ replaced by $W$: at stage $i$ for $T = T_{e(i)}$ from the current state, steer by a canonical shortest word $\sigma_i$ into a canonically chosen reachable terminal SCC $S$, choose in $S$ a directed simple cycle $C_i$ of maximal simple-cycle output rate with labels $(\alpha_i, \beta_i)$, set $p_i := \alpha_i^{n_i}$ and $L_i := |p_i|$, take $w_i \in (\{0,1\}^2)^{L_i}$ as the next $L_i$ unused symbols of $W$, and take a canonical shortest connector $c_i$ inside $S$ from the state after $w_i$ to the base of $C_i$; then define $r_i := \sigma_i\, w_i\, c_i$ and $Z = r_1 p_1\, r_2 p_2 \cdots$. Finally define $(X(t), Y(t)) = Z(t)$ coordinatewise. In particular, $|w_i| = |p_i| = L_i$ and $|\sigma_i|, |c_i| \le |Q| - 1$.

**Finite-state dimensions of $X$ and $Y$.** The computation $\dim_{\mathrm{FS}}(X) = \dim_{\mathrm{FS}}(Y) = \frac{1}{2}$ is the same as for the one-source sequence $X$ in Section 3, applied to each coordinate: the random-phase concatenation in each coordinate is obtained from consecutive blocks of the normal sequence $A$ (resp. $B$) by inserting paddings of length $|\sigma_i| + |c_i| = o(n_i)$, so it remains normal by Lemma 18; then the lower bound uses the same sliding-window mixture/concavity argument as Lemma 27 with mixing weight $\frac{1}{2} - o(1)$ (since $|w_i| = |p_i|$ and overhead is $o(L_i)$), and the upper bound uses the same heavy-atom argument as Lemma 30 from the long repetitions in the predictable phases.

**Bounding $\dim_{\mathrm{FS}}(T(X,Y))$.** Now fix any two-input transducer $T = T_e$. If $T(X,Y)$ is finite then $\dim_{\mathrm{FS}}(T(X,Y)) = 0$. Otherwise consider infinitely many stages $i$ with $e(i) = e$. At such a stage, $T$ begins reading $p_i = \alpha_i^{n_i}$ at the base of the chosen maximum-rate cycle,

hence the output on $p_i$ equals $b_i = \beta_i^{n_i}$. Let $s_i$ be the output produced on the SCC-internal suffix $w_i c_i$ of the same stage. Applying the rate bound (Lemma 20) inside the terminal SCC $S$,

$$|s_i| \;\le\; \rho_S(|w_i| + |c_i|) + O_T(1) \;=\; \rho_S|p_i| + o(|p_i|) + O_T(1) \;=\; |b_i| + o(|b_i|).$$

The steering word $\sigma_i$ contributes only $o(|b_i|)$ output (bounded output per input symbol), and all earlier stages contribute $o(|b_i|)$ output by the same fast-growth bookkeeping used in the single-source analysis. Hence $b_i$ occupies asymptotically half of the output prefix at the end of stage $i$. As in Lemma 34, this implies that for suitable $\ell = j|\beta_i|$ a single disjoint $\ell$-block appears with frequency at least $\frac{1}{2} - \varepsilon_i$ in infinitely many such output prefixes, where $\varepsilon_i \to 0$ along these stages; applying Lemma 17 yields

$$H_\ell(T(X,Y)) \le \tfrac{1}{2} + \varepsilon_i + \frac{1}{\ell},$$

and since $\varepsilon_i \to 0$ and $\ell$ can be taken arbitrarily large, it follows that $\dim_{\mathrm{FS}}(T(X,Y)) \le \frac{1}{2}$.

**Relative independence.** We prove $\dim_{\mathrm{FS}}^{Y}(X) = \dim_{\mathrm{FS}}(X) = \frac{1}{2}$; the other direction is symmetric. The upper bound $\dim_{\mathrm{FS}}^{Y}(X) \le \dim_{\mathrm{FS}}(X) = \frac{1}{2}$ is Lemma 15. For the matching lower bound, fix $\ell \ge 1$ and consider the *sliding* empirical distribution of $\ell$-blocks: for $t \ge 1$ let $U_t := X[t : t+\ell-1] \in \{0,1\}^\ell$ and $V_t := Y[t : t+\ell-1] \in \{0,1\}^\ell$. Now, for $N$ let $\widehat{P}_N$ be the empirical distribution of $(U_t, V_t)$ over $t \in \{1, \dots, N\}$. Write

$$\widehat{H}_\ell^{\mathrm{sl}}(X \mid Y; N) := \frac{1}{\ell}\, H_{\widehat{P}_N}(U \mid V), \qquad H_\ell^{\mathrm{sl}}(X \mid Y) := \liminf_{N\to\infty} \widehat{H}_\ell^{\mathrm{sl}}(X \mid Y; N).$$

By the equivalence recorded in the preliminaries (disjoint-block vs. sliding-window conditional $\ell$-block entropy rates yield the same $\dim_{\mathrm{FS}}^{Y}(X)$; see the remark following Definition 13), it suffices to show $H_\ell^{\mathrm{sl}}(X \mid Y) \ge \frac{1}{2}$ for every $\ell$.

Classify a starting position $t \in \{1, \dots, N\}$ as *source* if the paired $\ell$-window of $Z = \langle X, Y\rangle$ lies fully inside some source chunk $w_i$, *predictable* if it lies fully inside some $p_i$, and *boundary* otherwise. Exactly as in the fast-growth bookkeeping used throughout Section 3, boundary positions have vanishing density: the number of stage boundaries up to length $N$ is $O(\log\log N)$ and the total steering/connector length up to that point is $o(N)$, so only $o(N)$ starting positions yield windows intersecting boundaries or lying inside $\sigma_i, c_i$. Moreover, since $|w_i| = |p_i|$ and $|\sigma_i| + |c_i| = o(|w_i|)$, the source starting positions occupy a $\frac{1}{2} - o(1)$ fraction of $\{1, \dots, N\}$.

Restricting to source starting positions, the paired $\ell$-windows of $(X, Y)$ coincide with the paired $\ell$-windows of a long prefix of the source $W$ up to deletion of $o(N)$ windows (only those within $\ell - 1$ of the $w_i$–$w_{i+1}$ boundaries). Since $W$ is normal over $\{0,1\}^2$, the empirical *sliding* $\ell$-window distribution on $\{0,1\}^\ell \times \{0,1\}^\ell$ along source starting positions approaches uniform, and hence by Lemma 16 the conditional entropy per $\ell$-window of the $X$-block given the $Y$-block along source positions tends to 1. Let $U, V$ be distributed according to $\widehat{P}_N$. Then

$$\widehat{H}_\ell^{\mathrm{sl}}(X \mid Y; N) = \frac{1}{\ell} H(U \mid V) = \frac{1}{\ell} \sum_v \Pr[V = v]\; H(U \mid V = v).$$

For each $v$, the conditional distribution $U \mid (V = v)$ is a convex combination of the conditional distributions induced by source, predictable, and boundary starting positions among the times with $V = v$, and Shannon entropy is concave; thus $H(U \mid V = v)$ is at least the source weight (for that $v$) times the source-only conditional entropy. Averaging over $v$ yields

$$\widehat{H}_\ell^{\mathrm{sl}}(X \mid Y; N) \;\ge\; \Big(\frac{\#\{\text{source starts}\}}{N}\Big) \cdot (1 - o(1)).$$

Since the source fraction is $\geq \frac{1}{2} - o(1)$, taking $\liminf_{N\to\infty}$ gives $H_\ell^{\mathrm{sl}}(X \mid Y) \geq \frac{1}{2}$ for every $\ell$, and therefore $\dim_{\mathrm{FS}}^{Y}(X) \geq \frac{1}{2}$. Combining with the upper bound yields $\dim_{\mathrm{FS}}^{Y}(X) = \frac{1}{2}$, and symmetrically $\dim_{\mathrm{FS}}^{X}(Y) = \frac{1}{2}$. ◀

We remark that the above two-source construction extends to $k$ mutually relatively independent sources, and in fact to any prescribed rational finite-state dimension.

▶ **Theorem 42** ($k$ mutually relatively independent sources of rational dimension $s$)**.** *Fix $k \geq 2$ and a rational $s \in (0,1)$. There exist sequences $X_1, \ldots, X_k \in \{0,1\}^\infty$ such that, for each $i$,*

$$\dim_{\mathrm{FS}}(X_i) = \dim_{\mathrm{FS}}^{\widehat{X}_i}(X_i) = s, \qquad \widehat{X}_i = \langle X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k \rangle,$$

*and for every $k$-input finite-state transducer $T$, $\dim_{\mathrm{FS}}(T(X_1, \ldots, X_k)) \leq s$.*

To obtain this, write $s = \frac{a}{a+b}$ with $a, b \in \mathbb{N}$ and repeat the stage construction over input alphabet $\{0,1\}^k$. The two-source argument generalizes directly: replace the paired alphabet $\{0,1\}^2$ by $\{0,1\}^k$ and rebalance each stage so that the source and predictable portions have lengths in the ratio $a : b$. The remaining details are routine adaptations of the two-source proof (including the relative-independence verification with oracle $\widehat{X}_i$).

## 5 Discussion and open problems

This work highlights a contrast between two extraction landscapes. In the effective setting, suitable algorithmic independence between two positive-dimension sources can enable a uniform procedure to raise randomness rate arbitrarily close to 1 (e.g. Zimand's two-source phenomenon). By contrast, we show that, for *finite-state* dimension, even multiple independent streams (under either Kolmogorov complexity independence or finite-state relative independence) can fail to yield *any* guaranteed increase: there are sources of dimension $s$ from which *every* finite-state transducer outputs dimension at most $s$.

A natural way to organize the resulting picture is as a broader "threshold program": identify which computational resources first make multi-source extraction possible under meaningful independence hypotheses, and which resources remain too weak. Finite-state transduction is one extreme (single-pass, constant memory, streaming output), while Turing-computable reductions lie at the other. A particularly relevant next step is the pushdown level: pushdown transducers strictly extend finite-state transducers by allowing an unbounded stack while remaining online. It is therefore natural to ask whether stack memory suffices to resurrect a two-source extraction phenomenon under the same independence notions considered here.

An interesting open question concerns finite-state strong dimension. Our constructions force low-entropy behavior along infinitely many prefix lengths and therefore fundamentally exploit the $\liminf$ nature of $\dim_{\mathrm{FS}}$; it is not clear whether an analogous non-extraction phenomenon holds for $\mathrm{Dim}_{\mathrm{FS}}$, which is governed by $\limsup$ behavior. *For a given rational $s < 1$ (for example $s = \frac{1}{2}$), does there exist a sequence $X$ with $\mathrm{Dim}_{\mathrm{FS}}(X) = s$ such that every finite-state transducer $T$ satisfies $\mathrm{Dim}_{\mathrm{FS}}(T(X)) \leq s$?*

Another promising direction is to study extraction under the oracle-aided finite-state independence notion of Becher, Carton, and Heiber [2], defined via conditional finite-state compression by 2-deterministic 3-automata. In this model the compressor reads an input stream $x$ and an auxiliary stream $y$ on separate tapes and may advance one tape while reading $\varepsilon$ on the other, so oracle access can be inherently asynchronous; moreover, the conditional ratio $\rho(x/y)$ charges only symbols consumed from $x$ (oracle reads from $y$ are not counted). It is therefore natural to ask whether this oracle access model permits any finite-state dimension amplification from independent sources, or whether an analogue of our multi-source non-extraction phenomenon persists in this regime.

## References


1 Krishna Athreya, John M. Hitchcock, Jack H. Lutz, and Elvira Mayordomo. Effective strong dimension in algorithmic information and computational complexity. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 615–624, New York, NY, USA, 2007. ACM.

2 Verónica Becher, Olivier Carton, and Pablo Ariel Heiber. Finite-state independence. *Theory of Computing Systems*, 62(7):1555–1572, 2018. doi:10.1007/S00224-017-9821-6.

3 Chris Bourke, John M. Hitchcock, and N. V. Vinodchandran. Entropy rates and finite-state dimension. In *Proceedings of the 32nd International Colloquium on Automata, Languages, and Programming*, volume 3580 of *Lecture Notes in Computer Science*, pages 144–155, Berlin, 2005. Springer.

4 D. G. Champernowne. Construction of decimals normal in the scale of ten. *J. London Math. Soc.*, 2(8):254–260, 1933.

5 Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley Series in Telecommunications. John Wiley & Sons, Inc., New York, 1991.

6 Jin-Yi Dai, James I. Lathrop, Jack H. Lutz, and Elvira Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310(1-3):1–33, 2004. doi:10.1016/S0304-3975(03)00244-5.

7 David Doty. Dimension extractors and optimal decompression. *Theory of Computing Systems*, 43(3):425–463, 2008. doi:10.1007/S00224-007-9024-7.

8 Lance Fortnow, John M. Hitchcock, A. Pavan, N.V. Vinodchandran, and Fengming Wang. Extracting kolmogorov complexity with applications to dimension zero-one laws. *Information and Computation*, 209(4):627–636, 2011. doi:10.1016/j.ic.2010.09.006.

9 John M. Hitchcock. Fractal dimension and logarithmic loss unpredictability. *Theoretical Computer Science*, 304(1):431–441, 2003. doi:10.1016/S0304-3975(03)00138-5.

10 Richard M. Karp. A characterization of the minimum cycle mean in a digraph. *Discrete Mathematics*, 23(3):309–311, 1978. doi:10.1016/0012-365X(78)90011-0.

11 Anton Kozachinskiy and Alexander Shen. Automatic Kolmogorov complexity, normality, and finite-state dimension revisited. *Journal of Computer and System Sciences*, 2021. doi:10.1016/j.jcss.2020.12.003.

12 Ming Li, Paul Vitányi, et al. *An introduction to Kolmogorov complexity and its applications*, volume 3. Springer, 2008.

13 Jack H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003. doi:10.1016/S0890-5401(03)00187-1.

14 Jack H. Lutz and Elvira Mayordomo. Computing absolutely normal numbers in nearly linear time. *Information and Computation*, 281:104746, 2021. doi:10.1016/j.ic.2021.104746.

15 Elvira Mayordomo. A kolmogorov complexity characterization of constructive hausdorff dimension. *Information Processing Letters*, 84(1):1–3, 2002. doi:10.1016/S0020-0190(02)00343-5.

16 Joseph S. Miller. Extracting information is hard: A turing degree of non-integral effective hausdorff dimension. *Advances in Mathematics*, 226(1):373–384, 2011. doi:10.1016/j.aim.2010.06.024.

17 Satyadev Nandakumar, Subin Pulari, and Akhil S. Finite-state relative dimension, dimensions of a. p. subsequences and a finite-state van lambalgen's theorem. *Information and Computation*, 298:105156, 2024. doi:10.1016/j.ic.2024.105156.

18 Igal Sason. Entropy bounds for discrete random variables via maximal coupling. *IEEE Trans. Inf. Theor.*, 59(11):7118–7131, 2013. doi:10.1109/TIT.2013.2274515.

19 C.-P. Schnorr and H. Stimm. Endliche Automaten und Zufallsfolgen. *Acta Informat.*, 1(4):345–359, 1971/72. doi:10.1007/bf00289514.

20 Alexander Shen. Conditional normality and finite-state dimensions revisited. *Theoretical Computer Science*, 1057:115570, 2025. doi:10.1016/j.tcs.2025.115570.

21 Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972. doi:10.1137/0201010.

22 Marius Zimand. Two sources are better than one for increasing the kolmogorov complexity of infinite sequences. *Theory of Computing Systems*, 46(4):707–722, 2010. doi:10.1007/S00224-009-9214-6.